

Self-Organizing Feature Map With Improved Performance by Non-Monotonic Variation of the Learning Rate

Field of the Invention

5 The invention relates to Self-Organizing Feature Maps (SOFM) which are neural networks that transform an input of arbitrary dimension into a one or two dimensional discrete map subject to a topological (neighborhood preserving) constraint and more particularly to such SOFMs in which the
10 initial values of the weight vectors are random.

Background

Neural networks occupy a large branch of research in machine intelligence. Artificial neural networks are
15 information-processing devices inspired by the interconnected, parallel structure of animal brains. They take the form of software or hardware networks having collections of mathematical models that emulate some of the observed characteristics of nervous systems and analogies
20 to adaptive biological learning. Generally, they are composed of large numbers of interconnected processing elements, which can be realized in software or hardware, that are analogous to the neurons of an animal brain. The

connections between these processing elements are weighted in a fashion believed to be analogous to synapses.

Training a neural network involves making adjustments to the "synaptic" connections that exist between the
5 neurons (i.e., the values of the weights). Training is performed by exposure to a set of input/output data where the training algorithm iteratively adjusts the connection weights. These connection weights store the knowledge necessary to solve specific problems.

10 Neural networks are being applied to greater and greater numbers of practical problems, including very complex ones. They are particularly suited to pattern recognition and classification problems having many inputs such as speech recognition, character and signal
15 recognition, and functional prediction and system modeling where the physical processes are not understood or are highly complex.

There are many types of neural networks. Some of the more popular include the multilayer perceptron, which is
20 generally trained with the backpropagation of error algorithm, learning vector quantization, radial basis function, Hopfield, and SOFM. Some are classified as feedforward and others as recurrent (i.e., implement

feedback) depending on how data is processed through the network. Another feature of neural networks is the mechanism by which it is trained. Some use a technique called supervised training while others are referred to as
5 unsupervised or self-organizing. In supervised training, the network is guided by an instruction process while in unsupervised algorithms the data is clustered into similar groups based on the attributes that provide the inputs to the algorithms.

10 The SOFM or Kohonen artificial neural network is a type of unsupervised neural network. In unsupervised learning, an untrained neural network is exposed to examples or input vectors and internal parameters adjusted. In SOFMs, all the neurons of the network receive the same
15 input. The nodes engage in competition with their neighbors and at each stage of a self-learning process, the one with most activity "wins." Learning is based on the concept of winner neurons.

Unsupervised learning allows the objects to be grouped
20 together on the basis of their perceived closeness in n-dimensional hyperspace (where n is the number of variables or observations made on each object). Such methods, then, although in some sense quantitative, are better seen as

qualitative since their chief purpose is merely to distinguish objects or populations.

Referring to Fig. 1, SOFMs provide an objective way of classifying data through self-organizing networks of artificial neurons. There are two layers, an input layer 110 and a competition layer 100. Each node of the input layer may be connected (as indicated by connectors 120) to the entire set of nodes in the competition layer. In an example configuration, each neuron may be connected to its eight nearest neighbors on a grid. The neurons store a set of weights (a weight vector) each of which corresponds to one of the inputs in the data. The objective of a Kohonen network is to map input vectors (patterns) of arbitrary dimension N onto a discrete map lying in a competition layer of arbitrary dimension, but typically of 1 or 2 dimensions. The algorithm adjusts weights so that patterns close to one another in the input space should be close to one another in the map: they should be topologically ordered.

The learning process is as follows: First, the weights for each output unit are initialized, typically to random starting values. An iterative process is performed that end when weight changes are negligible. For each of a

number of input patterns a winning output node and all units in the neighborhood of the winner are identified and the weight vectors for all of these nodes updated. The winning output unit is simply the unit with the weight
5 vector that has the smallest Euclidean distance to the input pattern. The neighborhood of a unit is defined as all units within some distance of that unit on the map (not in weight space). If the size of the neighborhood is 1 then all units no more than 1, either horizontally or
10 vertically, from any unit fall within its neighborhood. The weights of every unit in the neighborhood of the winning unit (including the winning unit itself) are updated such that each unit in the neighborhood is moved closer to the input pattern. As the iterations are
15 performed, the learning rate is reduced. If the parameters are well chosen the final network should capture the natural clusters in the input data.

The factor that governs the size of the weight alterations is known as the learning rate. The adjustments
20 to each item in the weight vector are made in accordance with the following:

$$\Delta W_{i,j} = \alpha(I_j - W_{i,j}) \sin d / (2d)$$

where W_{ij} is the j^{th} weight of the i^{th} node, α is the learning rate, I_j is the j^{th} component of the input vector, and d is the distance between the current node and the winner. The above formula is one of a number that are known in the prior art and the invention to be discussed in subsequent section could make use of it or any other. As mentioned, the process of training continues until the changes in the weights falls below some predetermined value in successive iterations.

10 The effect of the "learning rule" (weight update algorithm) is to distribute the neurons evenly throughout the region of n-dimensional space populated by the training set. The neuron with the weight vector closest to a given input pattern will win for that pattern and for any other
15 input patterns that it is closest to. Input patterns which allow the same node to win are then deemed to be in the same group, and when a map of their relationship is drawn a line encloses them. In the resulting map, it is possible to examine closely relationships between the items in the
20 training set and visualize these relationships even for complex structures in high-dimensional input spaces.

There are two phases to the process of generating a solution. In an initial first stage, in the prior art, the

learning rate begins at a high value close to unity and is gradually monotonically decreased. The rate of decrease may be exponential, linear, or some other and according to the prior art, the particular pattern has not generally
5 been regarded as particularly important. During the initial phase, called the "ordering phase" the topological ordering of the weight vectors takes place. A long convergence phase follows and is associated with fine-tuning of the map. The learning rate is maintained at low
10 values (well below 0.1, for example) and monotonically and progressively decreased with each iteration.

The quality of a SOFM solution for any given problem is by no means assured. Where the input vectors are high-dimensional and complex, results may take a very long time
15 to converge and may even produce poor end results.

Summary of the Invention

In a SOFM, rather than reduce the learning rate monotonically during progression of the iteration process,
20 the learning rate is selected in a random or sporadic fashion. It has been found experimentally that this increases the rate of convergence in many instances. According to the prior art, the learning rate should be

time-varying, but decreased monotonically. However, it has been discovered that, at least in some contexts, particularly in the case of higher dimensional input spaces, sporadic variation of the learning rate, at least
5 during the initial stages, leads to improved performance.

The invention will be described in connection with certain preferred embodiments, with reference to the following illustrative figures so that it may be more fully understood. With reference to the figures, it is stressed
10 that the particulars shown are by way of example and for purposes of illustrative discussion of the preferred embodiments of the present invention only, and are presented in the cause of providing what is believed to be the most useful and readily understood description of the principles and conceptual aspects of the invention. In
15 this regard, no attempt is made to show structural details of the invention in more detail than is necessary for a fundamental understanding of the invention, the description taken with the drawings making apparent to those skilled in
20 the art how the several forms of the invention may be embodied in practice.

Brief Description of the Drawings

Fig. 1 is a diagrammatic representation of a self ordering feature map according to the prior art and consistent with embodiments of the invention.

5 Fig. 2 is a flow chart representing an algorithm for implementing the invention according to an embodiment thereof.

Fig. 3 is an illustration of selective random generation of learning rate parameters with progress of training of a SOFM.

Fig. 4 is an illustration of another manner of selective random generation of learning rate parameters with progress of training of a SOFM.

15 Detailed Description of the Preferred Embodiments

Referring to Fig. 2, an algorithm for implementing the invention begins with the initialization of the weight vectors in step S10. Random values for all of the weight vectors is a typical choice. In step S20, a sample input vector is drawn at random from a pool of input vectors being used for training the SOFM. The selected vector is applied to the input nodes and a winning competition layer

node is identified in step S30 according to a minimum Euclidean distance value:

$$D_i = \sqrt{\sum_{j=1}^n (I_j - W_{i,j})^2}$$

where $W_{i,j}$ is the j^{th} weight of the i^{th} node, I_j is the j^{th} component of the input vector, and D_i is the distance of the i^{th} node. The node with the lowest distance value is the winner. In step S40, a random value of the learning rate is generated. This learning rate is used to update the weights of the winning node and neighboring nodes.

10 In step S50, the winning node and all neighborhood nodes are then updated according to the following formula.

$$\Delta W_{i,j} = \alpha (I_j - W_{i,j}) \sin d / (2d)$$

where $W_{i,j}$ is the j^{th} weight of the i^{th} node, α is the random learning rate, I_j is the j^{th} component of the input vector, and d is the distance between the current node and the winner. As mentioned, this process of training continues until the changes in the weights falls below some predetermined value in successive iterations. The test is made in step S60.

20 Note that the above formulas are examples for purposes of describing an embodiment. There are other formulas that

may be used for updating the weights and the invention is not limited to any particular one.

Referring now to Fig. 3, the generation of successive values of the learning rate can be performed in various ways consistent with the present invention. Preferably, the values over which the random rates range should become smaller as the simulation progresses. In one example embodiment illustrated in Fig. 3, the learning rate is a random value between bounds 161 and 162 which decrease gradually as the training process progresses. This need not be a monotonic reduction in range, as illustrated at 160 in Fig. 3, but the learning rates preferably get smaller in value, and range over a smaller range, as the simulation progresses. In another embodiment, the learning rate is varied in a similar range 160, but is varied cyclically or pseudorandomly. Preferably, the learning rate varies around unity initially and falls to values near zero decreasing by several orders of magnitude.

Referring to Fig. 4, also, another alternative for generating may permit random variation (illustrated at 170) of the learning rate during the ordering phase of the training and switchover to monotonic reduction (illustrated at 180) of the learning rate for the convergence phase.

Although particular embodiments of the present invention have been shown and described, it will be understood that it is not intended to limit the invention to the preferred embodiments and it will be obvious to those skilled in the art that various changes and modifications may be made without departing from the spirit and scope of the present invention. Thus, the invention is intended to cover alternatives, modifications, and equivalents, which may be included within the spirit and scope of the invention.